

Misc. IT Sem-II 78:25 dt 07/01/2020

Paper / Subject Code: 82232 / Information Technology : Paper I - Big Data Analytics (R2020)

FA

Time: (2.30hrs)

[Total Marks: 75]

- N. B.:
- (1) All questions are compulsory.
 - (2) Make suitable assumptions wherever necessary and state the assumptions made.
 - (3) Answers to the same question must be written together.
 - (4) Numbers to the right indicate marks.
 - (5) Draw neat labelled diagrams wherever necessary.
 - (6) Use of Non-programmable calculators is allowed.

Q1. Attempt three of the following.

15 Marks

- a. List the challenges with big data?
- b. What is the role of support in apriori algorithm? Also explain how the Apriori property works with a neat diagram.
- c. Explain the use of decision tree, when we should have used it?
- d. What are the components of Spark Stack? Explain it.
- e. What are the Computing with Keys.

Q2. Attempt three of the following.

15 Marks

- a. List different phases of the Data Analytics Lifecycle? Explain each in detail.
- b. Write a short note on Diagnostics.
- c. Brief the ID3 Algorithm.
- d. What is tokenization? Explain how it is used in text analysis.
- e. Explain the concept of Hadoop Cluster.

Q3. Attempt three of the following.

15 marks

- a. Describe the Term Frequency-Inverse Document Frequency method.
- b. Explain the Box-Jenkins Methodology.
- c. Explain Hadoop Architecture in detail.
- d. Write short note on: A typical Spark Application
- e. How MapReduce can be implemented on a Cluster? Explain its all phases with a neat diagram.

Q4. Attempt three of the following.

15 marks

- a. Explain the steps involved in Text analysis with example.
- b. Explain the MapReduce Framework in details.
- c. Explain the following Advanced Map Reduce concepts:
a) Combiners b) Partitioners c) Job Chaining
- d. Write a short note on Hadoop Distributed File System.
- e. Demonstrate the process of Hadoop streaming in a MapReduce context.

Q5. Attempt three of the following.

15 marks

- a. How to categorize documents by topics? Explain in detail.
- b. What is sentiment analysis? How it can be carried out? Explain it in detail.
- c. Write a short note on : The Data Science Pipeline
- d. Explain the Spark Execution model
- e. Write a short note on Combiners in advanced MapReduce context.
